U.F.R. DE DROIT, ET DES SCIENCES SOCIALES
Année Universitaire **2019-2020**
Licence 3ème année, ÉCONOMIE
Econométrie 2
Nom des enseignants **:**
M. Alain Bousquet CM et Yann Kossi TD

1 Introduction

L'hétéroscédasticité se produit lorsque la variance pour toutes les observations dans un ensemble de données n'est pas la même. Dans cette application, vous examinerez les conséquences de l'hétéroscédasticité, vous trouverez des moyens de la détecter et vous verrez comment on peut corriger l'hétéroscdasticité en utilisant la régression avec des erreurs standard robustes et la régression des moindres carrés pondérés.

Pourquoi devons-nous nous soucier de l'hétéroscédasticité? Parce que c'est une violation de l'hypothèse des moindres carrés ordinaires selon laquelle $var(y_i) = var(e_i) = \sigma^2$. En présence d'hétéroscédasticité, il y a deux conséquences principales sur les estimateurs des moindres carrés :

1. L'estimateur des moindres carrés est toujours un estimateur linéaire et sans biais, mais il n'est plus le meilleur. Autrement dit, il existe un autre estimateur avec une variance plus petite.

2. Les écarts-types calculées pour les estimateurs des moindres carrés sont incorrects. Cela peut affecter les intervalles de confiance et les tests d'hypothèse qui utilisent ces écarts-types, ce qui pourrait conduire à des conclusions trompeuses.

La plupart des données sont probablement hétéroscédastiques. Cependant, on peut toujours utiliser les moindres carrés ordinaires sans corriger l'hétéroscédasticité parce que si la taille de l'échantillon est suffisamment grande, la variance de l'estimateur des moindres carrés peut toujours être suffisamment petite pour obtenir des estimations précises.

2 Documents et données fournies

2.1 Base de données

Analyse du budget consacré à l'alimentation par les ménages

Description
Données en coupe instantanée de from 1980
nombre d'observations : 23972
observation : ménages
pays : France
Nom du fichier BudgetFood
La base de donnée contient les variables suivantes:
wfood part en % de la dépense totale pour l'alimentation
totexp dépense totale du ménage
age age de la personne de référence dans le ménage
size taille du ménage

town	taille de la ville de 1 petites villes à 5 pour les plus grandes
sex	sexe de la personne de référence dans le ménage

2.2 scripts R et Markdown

budgetFood.R	script pour les estimations
DevoirEconometrieNom1Nom2.Rmd	script pour output de l'analyse au forma html

3 Questions

Le modèle de base consiste à estimer le modèle linéaire suivant

$$wfood_i = \beta_0 + \beta_1 \; totexp_i + \varepsilon_i \tag{1}$$

Le programme budgetFood.R vous permettra d'obtenir les résultats de l'estimation de ce modèle simple. Ce même programme contient par ailleurs une analyse approfondie de l'hétéroscédasticité en trois parties détection, tests puis correction par les moindres carrés généralisés faisables.

1. Généralisez le modèle simple initial en incluant les autres variables age, size, town et sex. **Attention**, dans les données les variables size et town sont des variables numériques, vous pouvez donc les utiliser directement. Cependant vous pourrez aussi transformer ces variables en variables catégorielles et tester ainsi les effets des variables dummies incorporées au modèle.

   L'âge n'a peut être pas un effet linéaire sur la part du budget alloué à l'alimentation, vous ajouterez la variable age^2 dans votre modèle ou sinon vous rendrez cette variable catégorielle (classe d'âge à définir par vos soins) puis inclure les variables dummies correspondante dans votre modèle.

2. Vous devrez chercher si certains des effets croisés entre les variables sont significatifs.

3. Le modèle de départ est un simple modèle linéaire, vous devrez tester des modèles alternatifs avec des spécifications différentes par exemple la forme log linéaire et décider si cette spécification donne, en un certain sens, de meilleurs résultats.

4. Pour chaque choix de modèle (liste des variables retenues et spécification linéaire ou en log) examinez la présence d'hétéroscédasticité et mettez en oeuvre en les adaptant les tests et procédures de correction de ce problème tels que présentés pour le modèle **??**.

4 Devoir

Par groupe de préférence de 2 étudiants et maximum de 3, **vous restituerez vos scripts R, Markdown et l'output html** faisant état de vos recherches, tests et analyses de différents modèles sur la base de données.

Le nom des fichiers transmis doit **comporter clairement les noms des étudiants** de chaque groupe.

En cas de difficulté et pour répondre à vos questions, exposez vos questions par email. Au besoin nous échangerons à distance par le moyen le plus adapté.

Ce travail sera évalué. Il est à rendre au cours de la journée du **vendredi 4 mars et avant minuit dernier délai** par simple envoi d'un email à, conjointement, alain.bousquet@univ-tours.fr et yann.kossi@univ-tours.fr